数理解析研究所講究録
1127 巻 2000 年 153-170



# On the simulation of some functionals of diffusion processes*

Arturo Kohatsu Higa
*Universitat Pompeu Fabra, Departament d'Economia,*
*Ramón Trias Fargas 25-27 08005-Barcelona, Spain*

Roger Pettersson†
*Department of Mathematical Statistics, Box 118,*
*Lund University, 221 00 Lund, Sweden*

**Abstract**

We study the numerical approximation of some functionals of diffusion processes. In particular we study their weak rate of convergence when these functionals are sufficiently smooth. We also give a variance reduction method for simulations of densities of the functionals.

*Key words:* stochastic differential equations, weak approximation, numerical analysis.
*Mathematics Subject Classification (1991)*: 60H99, 34B10, 34B15, 65Nxx

## 1 Introduction

Let $(\Omega, \mathcal{F}, P)$ be the standard Wiener space supporting a $k$-dimensional Wiener process. Let $X$ be the diffusion defined as the unique solution to the following stochastic differential equation

$$X_t = x + \int_0^t a(X_s)\mathrm{d}s + \sum_{j=0}^{k} \int_0^t b_j(X_s)\mathrm{d}W_s^j, \quad t \in [0,T]. \tag{1.1}$$

Here $x \in \mathbb{R}^d$. Also, $a : \mathbb{R}^d \to \mathbb{R}^d$ and $b = (b_1, ..., b_k) : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^k$ are Lipschitz functions. The above stochastic integral is the stochastic Itô's integral.

Let $F : \Omega \times L^2([0,T]; \mu) \to \mathbb{R}$ be an a.s. (in $\Omega$) Fréchet differentiable function where $\mu$ is a finite measure on $[0,T]$. In this article we will focus on the simulation error for $F(X)$ when the Euler-Maruyama scheme is used to approximate $X$. Many examples can be found in the literature when $F(X) = f(X_t)$ for some fixed $t$ and $f$. Recently, many

*The full version of this article will appear elsewhere

†Partially supported by the EU grant ref. ERBF MRX CT96 0075A.

applied problems require the analysis of functionals that depend on the whole path of the diffusion. Such are the cases of $F(X) = \int_0^T g(X_s)\mathrm{d}W^j(s)$ or $F(X) = \int_0^T g(X_s)\mathrm{d}\mu_s$ where $\mu$ is a finite measure, $d = 1$ and $g$ is a smooth function with polynomial growth at infinity. Another interesting example is $F(X) = \max_{s \leq T} X_s$. Although this last example will not satisfy the conditions in the analysis to be exposed here, it reveals the necessity of studying such functionals.

Here we approximate $X$ using the classical Euler-Maruyama scheme $\bar{X}$. To define it, let $\pi = \{0 = t_0, ..., t_N = T\}$ with $\|\pi\| = \max\{t_{i+1} - t_i;\ i = 0, ..., N-1\} \leq h$ for some $h > 0$. Also define $\eta(s) = \max\{t_i;\ t_i < s\}$ and $\eta_1(s) = \min\{t_i;\ t_i \geq s\}$. Then $\bar{X}$ is defined as the unique solution of the following equation:

$$\bar{X}_t = x + \int_0^t a(\bar{X}_{\eta(s)})\mathrm{d}s + \sum_{j=0}^{k} \int_0^t b_j(\bar{X}_{\eta(s)})\mathrm{d}W_s^j, \quad t \in [0, T]. \tag{1.2}$$

It is well known that $\bar{X}$ converges to $X$ in many different types of convergence and under various conditions. Estimates of the error are also available. Here we propose to simulate $F(X)$ in order to approximate quantities of the type $E(f(F(X)))$ where $f$ will first be a smooth function, then measurable with at most polynomial growth, and then a Dirac delta function.

We will analyze the total error of approximation in each case. That is, we will measure the quantity

$$E|E(f(F(X))) - \frac{1}{n}\sum_{i=1}^{n} f(F(\bar{X}^i))|. \tag{1.3}$$

Here $\bar{X}^i$, $i = 1, ..., n$ denotes $n$ independent copies of $\bar{X}$. Usually the above error is divided into two terms one of which one is called the weak approximation error (i.e., $|E(f(F(X))) - E(f(F(\bar{X})))|$) and the other the Monte Carlo error.

The weak approximation error is usually bounded by $Ch$ for a positive constant $C$ independent of the partition $\pi$ and $h$ but depends on the coefficients $a$, $b$ and the functions $F$ and $f$. We will show that the rate in this case is bounded by $Ch$ and that this result is satisfied even when $f$ is only measurable or a distribution function such as the delta function.

One of the problems that we will also address is the fact that as $f$ becomes more degenerate, the Monte Carlo error becomes bigger and eventually it goes to infinity as $n \to \infty$. That is, consider the variance of the estimate

$$\frac{1}{n}\sum_{i=1}^{n}\{f(F(\bar{X}^i)) - E(f(F(\bar{X})))\}. \tag{1.4}$$

A simple calculation gives an estimate of the variance of the type $n^{-1}\mathrm{var}[f(F(\bar{X}^i))]$. This variance will go to zero if $f$ has nice properties (e.g. $f$ bounded). In some interesting cases like when one approximates density functions one needs to use $f(x) = \phi(\frac{x-y}{h})$ with $h$ small. Here $\phi$ may be, for example, the density function of a $N(0,1)$ distribution. In this case the variance is bounded by $C(n\sqrt{h})^{-1}$. This property is well known in the literature of kernel density estimation which also implies the necessity of the tuning of the parameters $h$ and $n$ in order to achieve convergence of (1.3).

We propose here a variance reduction method that will allow the simulation of the densities without incurring in such a big error. We will actually show that through an appropriate change there exist simulatable random variables $Y^i$ such that $E(Y^i) = E(f(F(\bar{X}))) + O(h)$ and

$$\frac{1}{n}\sum_{i=1}^{n}\{Y^i - E(f(F(\bar{X})))\}$$

has a variance of the order $n^{-1}$ for $f = \phi(\frac{x-y}{h})$. This will bring the total error (weak approximation and Monte Carlo error) to be of the order $n^{-1/2} + h$.

This method can be applied in general to any approximation for irregular functions of processes where the Malliavin derivative properties are well understood.

The problem of simulation of these quantities arises naturally in a variety of fields when information about the density or distribution functions is required. In general the above methods are useful to estimate the kernel density functions which are the basic building blocks in order to construct solutions to parabolic partial differential equations and stochastic partial differential equations as well. Here, we mention briefly the case of Asian options in finance as a potential application but certainly many others are available.

One of the problems associated with the proposal we make here is that as it can be expected if one uses the scheme proposed here the constant that determine the rates of convergence become bigger. In some cases it seems that they become extremely big. Then in order to reduce to a maximum the error we propose some further variance reduction techniques that should help improve the behaviour of such constants. Through some computational examples we have shown that this can be achieved.

# 2 Preliminaries

In this section we introduce the main tools that we will use throughout the article. We start with some basic tools from Malliavin calculus that will be used throughout the text. For further reference, see [7]. Let $(\Omega, \mathcal{F}, P)$ be the canonical Wiener space which supports a $k$-dimensional Wiener process $W = (W^1, ..., W^k)$. We will also use $W_s^0 = s$ as an extension of the above notation.

On this space one defines the notion of stochastic derivative $D = (D^1, ..., D^k)$ on the space $\mathbb{D}^{1,2}$ which is a Banach space with norm denoted by $\|\cdot\|_{1,2}$. Analogously one defines $\mathbb{D}^{k,p}$ and its associated norm $\|\cdot\|_{k,p}$.

That is, let $C_b^\infty(\mathbb{R}^n)$ be the set of $C^\infty$ functions $f : \mathbb{R}^n \to \mathbb{R}$ which have bounded derivatives of all orders. $\mathcal{S}$ denotes the class of smooth functionals, i.e., a real random variable $F$ belongs to $\mathcal{S}$ if and only if there exist $t_1, ..., t_n \in [0, T]$ and $f(x_{11}, x_{21}, ..., x_{kn})$ in $C_b^\infty(\mathbb{R}^{kn})$ such that $F = f(W_{t_1}, ..., W_{t_n})$.

For $p \geq 2$, $\mathbb{D}^{1,p}$ designates the Banach space which is the completion of $\mathcal{S}$ with respect to the norm

$$\|F\|_{1,p} = (EF^p)^{1/p} + (E(\sum_{j=1}^{k} \int_0^T (D_s^j F)^2 ds)^{p/2})^{1/p},$$

where $DF$ is the derivative of the smooth functional $F$. That is,

$$D_t^j F = \sum_{i=1}^{n} \frac{\partial f}{\partial x^{ji}}(W_{t_1}, ..., W_{t_n}) 1_{[0,t_i]}(t).$$

Also, let for $\alpha \in \mathbb{N}$, $\mathbb{D}^{\alpha,\infty} = \cap_{p\geq 1}\mathbb{D}^{\alpha,p}$ and $\mathbb{D}^{\infty} = \cap_{p\geq 1} \cap_{\alpha \geq 1} \mathbb{D}^{\alpha,p}$.

The adjoint of the closed unbounded operator $D : \mathbb{D}^{1,2} \to L^2([0,T] \times \Omega)$ is usually denoted by $\delta$ and is called the Skorohod integral. Its domain can be characterized as the set of measurable processes $u \in L^2([0,T] \times \Omega)$ such that there exists a positive constant $C$ that may depend on $u$ such that

$$|E(\int_0^T D_t F u_t \mathrm{d}t)| \leq C\|F\|_2,$$

for all $F \in \mathbb{D}^{1,2}$. Then the Skorohod integral for $u$ an element of its domain, is the square integrable random variable determined by the duality relation

$$E(\delta(u)F) = E(\int_0^T D_t F u_t \mathrm{d}t), \tag{2.1}$$

for all $F \in \mathbb{D}^{1,2}$. The Skorohod integral turns out to be an extension of the classical Itô integral and it allows the integration of processes that are not necessarily adapted. $(\mathbb{D}^{\alpha,p})_{\mathrm{loc}}$ denotes the localization of $\mathbb{D}^{\alpha,p}$.

In order to avoid confusion we will use $D$ for the derivative defined above and $\nabla$ or the $'$ notation for classical derivatives of functions.

When considering densities of random variables we will use the concept of Malliavin covariance matrices. For this, define for $F \in (\mathbb{D}^{1,1})_{\mathrm{loc}}$ the Malliavin covariance matrix of $F$ as $\Delta_F^{ij} = \langle DF^i, DF^j \rangle_{L^2[0,T]}$. If $F \in \mathbb{D}^{\infty}$ and $\det \Delta_F^{-1} \in \cap_{p>1} L^p(\Omega)$, then $F$ has a smooth density.

An important component in the study of the density of $F$ is the integration by parts formula which can be established for any two random variables $F \in \mathbb{D}^{m+1,\infty}$, $G \in \mathbb{D}^{m,\infty}$, with $\Delta_F^{-1} \in \cap_{p>1} L^p$ and $f \in C_p^{\infty}$. Then the following formula holds:

$$E(f^{(m)}(F)G) = E(f(F)H_m(F,G)) \qquad \text{for} \quad m \geq 1,$$

where $H_m(F,G) = H(F, H_{m-1}(F,G))$ and

$$H_1(F,G) = H(F,G) = \delta(G\Delta_F^{-1}DF),$$

with $\delta$ defined as before.

Moreover, for any $p > 1$, there exist indices $p_1, p_2, p_3, \alpha_1, \alpha_2$, depending on $m$ and $p$ and a constant $C = C(m, p, p_1, p_2, p_3)$ such that

$$\|H_m(F,G)\|_p \leq C\|\Delta_F^{-1}\|_{p_1}^{\alpha_1} \|F\|_{m+1,p_2}^{\alpha_2} \|G\|_{m,p_3}. \tag{2.2}$$

Multi-dimensional formulae for the integration by parts are also available. See Nualart [7].

# 3 The Euler-Maruyama scheme

In this section we mention some results about the Euler-Maruyama scheme. Some preliminary results are mentioned in the following Lemma:

**Lemma 3.1** *Assume that $a, b \in C_b^r(\mathbb{R}^d)$. Then $X_t$, $\bar{X}_t \in (\mathbb{D}^{r,\infty})^d$ for all $t \in [0,T]$. Furthermore, there exists a finite positive constant $C_s$ independent of the partition such that*

$$\sup_{t_1,\ldots,t_m \in [0,T]} E(\sup_{t \leq T} |D_{t_1}^{j_1} \ldots D_{t_m}^{j_m} X_t|^p) \leq C_s,$$
$$\sup_{t_1,\ldots,t_m \in [0,T]} E(\sup_{t \leq T} |D_{t_1}^{j_1} \ldots D_{t_m}^{j_m} \bar{X}_t|^p) \leq C_s,$$
$$\sup_{t_1,\ldots,t_m \in [0,T]} E(\sup_{t \leq T} |D_{t_1}^{j_1} \ldots D_{t_m}^{j_m} (X_t - \bar{X}_t)|^p) \leq C_s h^{\frac{p}{2}}.$$

*Here $j_1, \ldots, j_m \in \{1, \ldots, k\}$ and $m \leq r$.*

The proof of this lemma is already classical. One can find close versions of this Lemma in many articles related to the numerical analysis of diffusions. For example, see Hu and Watanabe [3]. The method of proof here is exactly the same.

In the next Lemma we consider a useful expression related to the difference between the diffusion and its approximation. From its statement it becomes clear that this difference has an error that is determined by the differences $t_{i+1} - t_i$ and $W_{t_{i+1}} - W_{t_i}$.

**Lemma 3.2** *Let $a \in C_b^r(\mathbb{R}^d)$ and $b \in C_b^r(\mathbb{R}^d; \mathbb{R}^d \times \mathbb{R}^k)$, for some $r \geq 1$. Then we have*

$$X_t - \bar{X}_t = \sum_{j_1, j_2 = 0}^{k} \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1} A_s^{j_1, j_2} (W_s^{j_1} - W_{\eta(s)}^{j_1}) \mathrm{d}W_s^{j_2}. \tag{3.1}$$

*Here $\mathcal{E}$ is the unique $d \times d$ matrix valued adapted stochastic process solution of an appropriate linear stochastic differential equation with bounded coefficients. Furthermore, $A^{j_1, j_2}$ is an adapted stochastic process taking values in $\mathbb{R}^d$ such that $A_s^{j_1, j_2} \in \mathbb{D}^{r-1,\infty}$ for all $s \leq T$ and $m \leq r-1$,*

$$\sup_{t_1,\ldots,t_m \in [0,T]} E(\sup_{s \leq T} |D_{t_1}^{l_1} \ldots D_{t_m}^{l_m} A_s^{j_1, j_2}|^p) \leq C,$$

*for a positive constant $C$ independent of the partition, $l_1, \ldots, l_m \in \{1, \ldots, k\}$, $j_1$ and $j_2 \in \{0, \ldots, k\}$.*

*Proof:* First, using (1.1) and (1.2) we have

$$\begin{aligned} X_t - \bar{X}_t &= \int_0^t a'(\xi_s^0)(X_s - \bar{X}_s) ds + \int_0^t \nabla b^{\cdot, j}(\xi_s^j)(X_s - \bar{X}_s) dW_s^j \\ &+ \int_0^t a(\bar{X}_s) - a(\bar{X}_{\eta(s)}) ds + \int_0^t b(\bar{X}_s) - b(\bar{X}_{\eta(s)}) dW_s. \end{aligned} \tag{3.2}$$

Here we are using the multiple index summation notation. Also $\xi_s^0$ and $\xi_s^j$ are random points in the interval determined by $X_s$ and $\bar{X}_s$. In particular we will always understand the expression $a'(\xi_s^0)$ in its integral form. That is,

$$a'(\xi_s^0) = \int_0^1 a'(\bar{X}_s + \lambda(X_s - \bar{X}_s))d\lambda.$$

Similar remarks are assumed for all the random points that appear in the rest of this article.

Now, going back to Equation (3.2), we note that the equation is linear in $X - \bar{X}$. Therefore, if we define $\mathcal{E}$ as the unique solution to

$$\mathcal{E}_t = I + \int_0^t \nabla b^{\cdot,j}(\xi_s^j)\mathcal{E}_s dW_s^j,$$

and use the general formula for the solution of a linear stochastic differential equation, one has:

$$\begin{aligned} X_t - \bar{X}_t &= \mathcal{E}_t \int_0^t \mathcal{E}_s^{-1} \nabla b^{\cdot,j}(\epsilon_s^j) b(\bar{X}_{\eta(s)})(W_s - W_{\eta(s)})dW_s^j \\ &\quad -\mathcal{E}_t \int_0^t \mathcal{E}_s^{-1} \nabla b^{\cdot,j}(\xi_s^j) \nabla b^{\cdot,j}(\epsilon_s^j) b(\bar{X}_{\eta(s)})(W_s - W_{\eta(s)})d < W^j, W^j >_s . \end{aligned}$$

Here

$$\nabla b^{\cdot,j}(\epsilon_s^j) = \int_0^1 \nabla b^{\cdot,j}(\bar{X}_{\eta(s)} + \lambda(\bar{X}_s - \bar{X}_{\eta(s)}))d\lambda,$$

with the notation $\nabla b^{\cdot,0} = a'$. From here one obtains the statement of the Lemma. In particular one can identify exactly all the terms $A^{j_1,j_2}$. □

# 4 Smooth functionals of X

Here we study the weak approximation of smooth functionals of the diffusion $X$. For this we assume troughout this section that $F$ is Fréchet differentiable with a continuous derivative. More exactly we define $FC^r$ as the space of functionals $F : L^2([0,T]; \mathbb{R}^d; \mu) \to \mathbb{R}^l$ such that $F$ is continuously Fréchet differentiable $r$ times and its derivatives $\nabla_{t_1,\ldots,t_s} F(X) \in \mathbb{R}^d \times \mathbb{R}^l$ satisfy for some appropriate positive constants $C$ and $p$

$$|\nabla_{t_1,\ldots,t_s} F(X)| \leq C(1 + \sup_{u \leq T} |X_u|^p) \tag{4.1}$$

for almost all $t_1, \ldots, t_s \in [0,T]$ and $s \leq q$. We denote by $\nabla^{p_1} F^{p_2}(X)$ the $(p_1, p_2)$ element of the matrix $\nabla F(X)$ for $p_1 = 1, \ldots, d$ and $p_2 = 1, \ldots, l$.

**Lemma 4.1** *Let $F(\omega, \cdot) \in FC^r$ for almost all $\omega$ such that $\nabla_{t_1,\ldots,t_s} F(x) \in \mathbb{D}^\infty$ for all $s \leq r$ and $x \in L^2([a,b], \mu)$. Also assume that $a$, $b_j \in C_b^r(\mathbb{R}^d)$ for $j = 1, \ldots, k$. Then $\nabla_{t_1,\ldots,t_s} F(X) \in \mathbb{D}^\infty$. Furthermore,*

$$DF(X) = DF(\cdot)(X) + \int_0^T F_s'(X) DX_s \mathrm{d}\mu_s.$$

The proof of this statement uses classical techniques of chain rule formulae such as Proposition 1.2.2 in Nualart [7]. From now on, by a slight abuse of notation, we will say $F \in FC^r$ to mean that $F(\omega, \cdot) \in FC^r$ for almost all $\omega$.

**Theorem 4.2** *Let $F$ be an element of $FC^{r+1}$ and assume that $a$, $b_j \in C_b^{r+1}(\mathbb{R}^d)$ for $j = 1, ..., k$. Then*

$$\|F(X) - F(\bar{X})\|_{r,p} \leq C\sqrt{h},$$

*for a positive constant $C$ independent of $h$ and the partition $\pi$.*

Before introducing the next result we need to define some spaces. Let $C_p^r(\mathbb{R}^l)$ denote the space of $r$-times continuously differentiable functions such that their derivatives have polynomial growth at infinity.

**Theorem 4.3** *Assume that $a \in C_b^4(\mathbb{R}^d)$ and $b \in C_b^4(\mathbb{R}^d; \mathbb{R}^d \times \mathbb{R}^k)$, $f \in C_p^6(\mathbb{R}^l)$ and that $F$ is in $FC^6$. Then there exist constants $C_1$ and $R_h$ such that*

$$E(f(F(X))) - E(f(F(\bar{X}))) = C_1 h + R_h h^2,$$

*where $C_1$ is independent of $h$ and the partition. $R_h$ is a constant that depends on $h$ and satisfies* $\sup_{0<h<1} |R_h| < \infty$.

*Proof:* Applying the mean value theorem we have that

$$f(F(X)) - f(F(\bar{X})) = \int_0^1 f'(F(\bar{X}+\lambda(X-\bar{X})) \int_0^T \nabla_s F(\bar{X}+\lambda(X-\bar{X}))(X_s - \bar{X}_s) d\mu_s d\lambda.$$

Now we use Lemma 3.2 to obtain that

$$\begin{aligned} &E[f(F(X)) - f(F(\bar{X}))] \\ &= \int_0^1 \int_0^T E\Big[f'(F(\bar{X} + \lambda(X - \bar{X}))\nabla_s F(\bar{X} + \lambda(X - \bar{X})) \\ &\quad \sum_{j_1,j_2=0}^{k} \mathcal{E}_s \int_0^s \mathcal{E}_u^{-1} A_u^{j_1,j_2} (W_u^{j_1} - W_{\eta(u)}^{j_1}) dW_u^{j_2}\Big] d\mu_s d\lambda. \end{aligned} \tag{4.2}$$

Using the integration by parts formula (2.1) one obtains that

$$\begin{aligned} &E[f(F(X)) - f(F(\bar{X}))] \\ &= \int_0^1 \int_0^T \int_0^s \int_{\eta(u)}^u \sum_{j_1,j_2=0}^{k} E\Big[D_v^{j_1}\Big\{D_u^{j_2}\Big\{f'(F(\bar{X} + \lambda(X - \bar{X})) \\ &\quad \nabla_s F(\bar{X} + \lambda(X - \bar{X}))\mathcal{E}_s\Big\}\mathcal{E}_u^{-1} A_u^{j_1,j_2}\Big\}\Big] dv du d\mu_s d\lambda \end{aligned} \tag{4.3}$$

Proving that the integrand is uniformly bounded gives as a result

$$|E[f(F(X)) - f(F(\bar{X}))]| \leq Ch.$$

To finish the proof one replaces the integrand in (4.3) by

$$E\left[D_v^{j_1}\left\{D_u^{j_2}\left\{f'(F(X))\nabla_s F(X)E_s\right\}E_u^{-1}\hat{A}_u^{j_1,j_2}\right\}\right],$$

where $s \mapsto E_s$ is the first derivative of the flow associated with $X$. $\hat{A}$ is defined as $A$ using $X$ instead of $\bar{X}$ and $s$ instead of $\eta(s)$ in its definition. Then has to estimate $R_h$ through the estimation of the difference:

$$E\Bigg[D_v^{j_1}\left\{D_u^{j_2}\left\{f'(F(\bar{X}+\lambda(X-\bar{X}))\nabla_s F(\bar{X}+\lambda(X-\bar{X}))\mathcal{E}_s\right\}\mathcal{E}_u^{-1}A_u^{j_1,j_2}\right\}$$
$$- \; D_v^{j_1}\left\{D_u^{j_2}\left\{f'(F(X))\nabla_s F(X)E_s\right\}E_u^{-1}\hat{A}_u^{j_1,j_2}\right\}\Bigg],$$

and has to again apply the same procedure for each term in (4.2). □

# 5 Approximations for irregular functions

In what follows we will assume a condition that assures the existence and smoothness of the density of the random variable $F(X)$.
(H) $\det(\Delta_{F(X)})^{-1} \in \cap_{p>1} L_p(\Omega)$.

This condition assures that the random variable $F(X)$ has a smooth density. $\phi_r$ will represent the zero mean normal density with standard deviation $r$ and $\Phi_r$ the corresponding cumulative distribution function. We start with a preliminary Lemma.

**Lemma 5.1** *Assume (H). Then*

$$\sup_{h\in(0,1]} \|\det(\Delta_{F(\bar{X})+h^\beta \bar{W}_T})^{-1}\|_p < \infty.$$

*Proof:* The proof of the above lemma is obtained using the following 3 facts: Theorem 4.2, hypothesis (H) and Chebyshev's inequality. We will just sketch it here. Define the set $A := [|\det(\Delta_{F(\bar{X})}) - \det(\Delta_{F(X)})| < \frac{1}{2}\det(\Delta_{F(X)})]$. Then we have

$$\begin{aligned}
E(\det(\Delta_{F(\bar{X})+h^\beta\bar{W}_T})^{-p};A) &\le C_p E(\det(\Delta_{F(X)})^{-p}) \\
E(\det(\Delta_{F(\bar{X})+h^\beta\bar{W}_T})^{-p};A^c) &\le C_p(h^{2\beta}T)^{-p}P(A^c) \\
&\le C_p(h^{2\beta}T)^{-p}E(|\det(\Delta_{F(\bar{X})}) - \det(\Delta_{F(X)})|^{2m})^{1/2} \\
&\quad \times E(|\det(\Delta_{F(X)})|^{-2m})^{1/2}2^m \\
&\le C_p(h^{2\beta}T)^{-p}h^{m/2}E(|\det(\Delta_{F(X)})|^{-2m})^{1/2}2^m.
\end{aligned}$$

Then choosing $m$ big enough the result follows. □

**Proposition 5.2** *Assume that $a \in C_b^7(\mathbb{R}^d)$ and $b \in C_b^7(\mathbb{R}^d;\mathbb{R}^d\times\mathbb{R}^k)$. Also suppose that $F \in FC^7$ and that (H) is satisfied. Let $f:\mathbb{R}^l \to \mathbb{R}^m$ be a measurable function such that*

*$|f(x)| \leq C(1+|x|^r)$ for some positive constant $C$ and $r \geq 0$. If $\bar{W}$ denotes a Wiener process independent of $W$ then one has that for $\beta \geq 1/2$,*

$$E(f(F(X))) - E(f(F(\bar{X}) + h^\beta \bar{W}_T)) = C_1 h + R_h h^2,$$

*where $C_1$ is a positive constant independent of the partition and $h$. $R_h$ satisfies $\sup_{0<h<1} |R_h| < \infty$.*

Bally and Talay have obtained this result in the case $F(X) = X_t$ for fixed $t$. They require $f$ to be bounded and a uniform Hörmander type condition. This condition is weakened in our result.

Note that for (H) to be satisfied in this particular case we only need a Hörmander type condition to be satisfied at the initial point of the diffusion. This implication is proved in Kusuoka and Stroock [6].

*Proof of Proposition 5.2:* The proof consists of first proving the following approximation result

$$E(f(F(X))) - E(f(F(X) + h^\beta \bar{W}_T)) = C_1 h + R_h h^2.$$

This is done using the same arguments as in e.g. Lemma 3.4 of [4]. Informally it goes as follows. For $a > 0$ consider the Taylor expansion of (here the symbol $*$ denotes the convolution):

$$\begin{aligned} & E(f * \phi_a(F(X)) - f * \phi_a(F(X) + h^\beta \bar{W}_T)) \\ = \;& E((f*\phi_a)''(F(X)))h^{2\beta}T + \int_0^1 E((f*\phi_a)^{(4)}(F(X) + \lambda h^\beta \bar{W}_T)\bar{W}_T^4)h^{4\beta} d\lambda \\ = \;& C_1^a h + R_h^a h^2. \end{aligned}$$

Then $C_1^a$ and $R_h^a$ is rewritten using the integration by parts formula. The terms are bounded due to (2.2) and the hypothesis (H). From here one takes limits with respect to $a \to 0$.

Now to deal with the second term

$$E(f(F(X) + h^\beta \bar{W}_T) - f(F(\bar{X}) + h^\beta \bar{W}_T))$$

one uses the same proof as in Theorem 4.3. The difference rests that at the end one has to apply the Lemma 5.1 and the bounds (2.2). □

The above proposition shows that even if the functions are non-smooth the approximation method works well. What makes things not so optimistic is the Monte Carlo error which is characterized by the quantity $n^{-1/2}\mathrm{Var}(f(F(\bar{X}) + \sqrt{h}\bar{W}_T))$ which obviously behaves badly as $f$ becomes more degenerate. One such a case is when $f$ is the Dirac delta function. This case is studied in the next section and later we will address the issue of the Monte Carlo approximation.

Due to hypothesis (H), we know that the density of $F(X)$ exists and is smooth. In order to make the exposition of ideas simple we assume from now on that $l = 1$. One can use similar steps of the previous lemma to obtain an approximation theorem for the densities.

**Proposition 5.3** *Assume that $a \in C_b^8(\mathbb{R}^d)$ and $b \in C_b^8(\mathbb{R}^d; \mathbb{R}^d \times \mathbb{R}^k)$. Also suppose that $F \in FC^8$ and that (H) is satisfied. If $\bar{W}$ denotes a Wiener process independent of $W$ then one has that for $\gamma,\ \beta \geq 1/2$,*

$$E(\delta_x(F(X))) - E(\phi_{h^\gamma}(F(\bar{X}) + h^\beta \bar{W}_T - x)) = C_1(x)h + R_h(x)h^2.$$

*Here $C_1(x)$ is a positive constant independent of the partition and $h$ which is uniformly bounded in $x$. $R_h$ satisfies $\sup_{0<h<1, x\in\mathbb{R}^d} |R_h(x)| < \infty$.*

Here Bally and Talay required a uniform Hörmander condition together with a distance condition between $X_0$ and $x$. Here one realizes that such stringent conditions are not necessary. Hu and Watanabe obtained slower rates of convergence.

*Proof:* The proof is as in the proof of Proposition 5.2. Except that one has to apply integration by parts once more so that one obtains a bounded function instead of an unbounded one. □

As a Corollary we obtain the following rate of convergence for the Monte Carlo method.

**Corollary 5.4** *Under the same conditions as above we have,*

$$E|E(\delta_x(F(X))) - \frac{1}{n}\sum_{i=1}^{n} \phi_{h^\gamma}(F(\bar{X}^i) + h^\beta \bar{W}_T^i - x)| \leq C_2(x)(h + \frac{1}{h^{\gamma/2}\sqrt{n}}).$$

*Here $C_2(x)$ is a positive constant independent of the partition and $h$, and uniformly bounded in $x$.*

This result reveals the necessity of "tuning" the value of the parameters $n$ and $h$ so that the resulting estimate is good enough. This is a well known result in non-parametric density estimation theory. The rate that one obtains in such a general theory is worse than the one obtained here.

In the next section we show that the rate obtained in the previous corollary can in fact be improved.

# 6 A Variance Reduction Method for irregular functions

In this section we propose a variance reduction method in order to reduce the Monte Carlo error characterized by $\mathrm{Var}(f(F(\bar{X}) + h^\beta \bar{W}_T)))$. To simplify the discussion we will assume the hypothesis (H) and that the coefficients are smooth with bounded derivatives. We start proving that in the particular case that $f = \phi_{h^\gamma}$ this variance in fact explodes.

**Lemma 6.1** *Assume the same hypothesis as in the Proposition 5.3. Then for $\gamma,\ \beta \geq 1/2$,*

$$h^\gamma \mathrm{Var}(\phi_{h^\gamma}(F(\bar{X}) + h^\beta \bar{W}_T - x)) \to E(\delta_x(F(X))) \text{ as } h \to 0.$$

In particular the variance will converge to $\infty$ if the density of $F(X)$ is not zero at $x$.
*Proof:* One has that

$$E(\phi_{h^\gamma}(F(\bar{X})+h^\beta\bar{W}_T-x))\to E(\delta_x(F(X)))\neq 0,$$

due to Proposition 5.3. Therefore it is enough to note that

$$E(\phi_{h^\gamma}(F(\bar{X})+h^\beta\bar{W}_T-x)^2)=\frac{1}{2\sqrt{\pi}h^\gamma}E(\phi_{h^\gamma/\sqrt{2}}(F(\bar{X})+h^\beta\bar{W}_T-x)).$$

□

The above result is also true for $\gamma$, $\beta>0$ but the rates of convergence or divergence are different. Also from Nualart [8, Proposition 4.2.2] one has criteria to determine the positivity of the density at a given point.

Obviously we have left aside the consideration of the number of independent copies being used for the Monte Carlo simulation. If ones tunes $n$ with $h$ one can obtain an estimation of the probability density at $x$. However, it is obvious that as $h$ is chosen small, $n$ will have to be bigger. Our proposal for variance reduction is to use the integration by parts formula in order to avoid this problem. That is, instead of considering the quantity

$$E(\phi_{h^\gamma}(F(\bar{X})+h^\beta\bar{W}_T-x))$$

we will use

$$E(\Phi_{h^\gamma}(F(\bar{X})+h^\beta\bar{W}_T-x)H(F(\bar{X})+h^\beta\bar{W}_T,1)) \tag{6.1}$$

and eventually we will propose something closer to

$$E(1(F(\bar{X})\geq x)H(F(\bar{X})+h^\beta\bar{W}_T,1)),$$

with $\beta\geq 2$.

**Theorem 6.2** *Assume that $a\in C_b^9(\mathbb{R}^d)$ and $b\in C_b^9(\mathbb{R}^d;\mathbb{R}^d\times\mathbb{R}^k)$. Also suppose that $F\in FC^9$ and that (H) is satisfied. Then there exists a simulatable random variable $Y$ such that*

$$\sup_{x\in\mathbb{R}}E|E(\delta_x(F(X)))-\frac{1}{n}\sum_{i=1}^{n}Y^i|\leq C(h+\frac{1}{\sqrt{n}}),$$

*where the constant $C$ is independent of the partition, $h$ and $n$. $Y^i$ denotes for $i=1,...,n$, $n$ independent copies of $Y$.*

We start with a series of Lemmas. The first proves that the quantity (6.1) gives the same approximation result as in Proposition 5.3.

**Lemma 6.3** *Let $\gamma$, $\beta\geq 1/2$. Then*

$$E(\delta_x(F(X)))-E(\Phi_{h^\gamma}(F(\bar{X})+h^\beta\bar{W}_T-x)H(F(\bar{X})+h^\beta\bar{W}_T,1))=C_1(x)h+R_h^1(x)h^2,$$

*where $C_1(x)$ is a positive constant independent of the partition and $h$, and is uniformly bounded in $x$. $R_h^1$ satisfies $\sup_{0<h<1,x\in\mathbb{R}^d}|R_h^1(x)|<\infty$. Furthermore,*

$$\mathrm{Var}(1(F(\bar{X})\geq x)H(F(\bar{X})+h^\beta\bar{W}_T,1))\leq C,$$

*where $C$ is a positive constant that is independent of $h$ and $x$.*

*Proof:* The first part of the statement is trivial. To prove the second statement it is enough to note that

$$\|H(F(\bar{X})+h^\beta\bar{W}_T,1)\|^2_{L^2(\Omega)} \leq C.$$

This follows from Lemma 5.1 and the bounds (2.2). □

**Lemma 6.4** *Let $\gamma \geq 1/2$ and $\beta > 4$. Then*

$$\begin{aligned}
&|E(\Phi_{h^\gamma}(F(\bar{X})+h^\beta\bar{W}_T-x)H(F(\bar{X})+h^\beta\bar{W}_T,1))\\
&\quad - E(1(F(\bar{X})\geq x)H(F(\bar{X})+h^\beta\bar{W}_T,1))|\\
&= C_2(x)h+R^2_h(x).
\end{aligned}$$

*Here $C_2(x)$ is a positive constant independent of the partition and $h$, and is uniformly bounded in $x$. $R^2_h$ satisfies $\sup_{0<h<1,x\in\mathbb{R}^d}|R^2_h(x)|<\infty$.*

*In the particular case that*
*Condition (H1): $\det(\Delta_{F(\bar{X})})^{-1}\in\cap_{p>1}L_p(\Omega)$.*
*Then one can replace the approximating expression by $E(1(F(\bar{X})\geq x)H(F(\bar{X}),1))$ and only needs $\gamma\geq 1/2$ and $\beta\geq 1/2$ for the same result to be satisfied.*

*Proof:* First we consider

$$\begin{aligned}
&E(1(F(\bar{X})+h^\beta\bar{W}_T h^\gamma\hat{W}_1\geq x)H(F(\bar{X})+h^\beta\bar{W}_T,1))\\
&-E(1(F(\bar{X})+h^\beta\bar{W}_T\geq x)H(F(\bar{X})+h^\beta\bar{W}_T,1))\\
&=D_2(x)h^{2\gamma}+S^2_h(x)h^{4\gamma}.
\end{aligned}$$

Here $D$ and $S$ satisfy the same properties as $C_2$ and $R^2$. This is obtained by using the same Taylor expansion and integration by parts techniques as in Proposition 5.2.

Secondly one proves

$$|E(1(F(\bar{X})+h^\beta\bar{W}_T\geq x)H(F(\bar{X})+h^\beta\bar{W}_T,1))-E(1(F(\bar{X})\geq x)H(F(\bar{X})+h^\beta\bar{W}_T,1))|\leq Ch^{\beta-\epsilon}$$

To prove this one considers the same set $A$ of the proof of Lemma 5.1. Then one defines $\varphi_A$ a "smooth version " of the set $A$. Then we have that is enough to consider

$$E\left[|1(F(\bar{X})+h^\beta\bar{W}_T\geq x)-1(F(\bar{X})\geq x)|^2(\varphi_A+(1-\varphi_A))\right]$$

The expectation restricted to $A^c$ or $1-\varphi_A$ can be treated as in Lemma 5.1. For the other one decomposes it in the probabilities

$$\begin{aligned}
&P(\{\omega\in\Omega; x>F(\bar{X})\geq -h^\beta\bar{W}_T+x\}\cap A)\\
&P(\{\omega\in\Omega; x-h^{\gamma/2}\bar{W}_T>F(\bar{X})\geq x\}\cap A).
\end{aligned}$$

Both probabilities are treated similarly so we will do the first.

$$P(\{\omega\in\Omega; x>F(\bar{X})\geq -h^\beta\bar{W}_T+x\}\cap A) = \int_{-\infty}^{x}(1-\Phi_{h^\beta}(x-z)E(\delta_z(F(\bar{X}))\varphi_A)dz$$

To finish one realizes that $(1-\Phi_{h^\beta}(x-z)$ is small than any polynomial of $h$ if $x-Ch^{\beta-\epsilon}>z$ for any $\beta>\epsilon>0$. Furthermore the term $E(\delta_z(F(\bar{X}))\varphi_A)$ is uniformly bounded. From these two facts the proof of the first case follows.

In the particular case that the Malliavin covariance matrix of $F(\bar{X})$ is non-degenerate the proof goes as above but the last step is not needed. □

A somewhat similar result can be achieved playing with the above techniques differently.

**Corollary 6.5** *Let $\gamma \geq 4$ and $\beta \geq 2$, then*

$$
\begin{aligned}
|E(\Phi_{h^\gamma}(F(\bar{X})+h^\beta \bar{W}_T-x)H(F(\bar{X})+h^\beta \bar{W}_T,1)) \quad & -E(1(F(\bar{X})\geq x)H(F(\bar{X})+h^\beta \bar{W}_T,1))| \\
& \leq C_2(x)h^2.
\end{aligned}
$$

*Here $C_2(x)$ is a positive constant independent of the partition and $h$ which is uniformly bounded in $x$.*

These two results show that one can play around with the construction of $Y$ according to the situation at hand. That is if $F(\bar{X})$ is highly degenerate or not.

*Proof of Theorem 6.2:* In order to show how to construct the random variables $Y$ one has to give efficient ways to approximate $H(F(\bar{X})+h^\beta \bar{W}_T,1)$. For this reason one has to give the following formulas that provide approximations for $D^j\bar{X}$ for $j=1,2$.

Note that from (1.2) one has for $l,m=1,...,k$,

$$
D_u^l\bar{X}_t = b_l(\bar{X}_{\eta(u)}) + \int_{\eta_1(u)}^t a'(\bar{X}_{\eta(s)})D_u^l\bar{X}_{\eta(s)}\mathrm{d}s + \int_{\eta_1(u)}^t b_j'(\bar{X}_{\eta(s)})D_u^l\bar{X}_{\eta(s)}\mathrm{d}W_s^j
$$

and

$$
\begin{aligned}
D_v^m D_u^l \bar{X}_t &= b_l'(\bar{X}_{\eta(u)})D_v^m\bar{X}_{\eta(u)} \\
&+\int_{\eta_1(u)}^t a''(\bar{X}_{\eta(s)})D_u^l\bar{X}_{\eta(s)}D_v^m\bar{X}_{\eta(s)} + a'(\bar{X}_{\eta(s)})D_v^m D_u^l\bar{X}_{\eta(s)}\mathrm{d}s \\
&+\int_{\eta_1(u)}^t b_j''(\bar{X}_{\eta(s)})D_u^l\bar{X}_{\eta(s)}D_v^m\bar{X}_{\eta(s)} + b_j'(\bar{X}_{\eta(s)})D_v^m D_u^l\bar{X}_{\eta(s)}\mathrm{d}W_s^j.
\end{aligned}
$$

These equations can be also written in recursive form to allow their simulation based only on the simulation of the increments of the Wiener process. Also note that $D_u\bar{X}_t$ as well as $D_uD_v\bar{X}_t$ are constant for $u,v\in(t_i,t_{i+1}]$.

Furthermore one can write an explicit expression for $H$ as follows:

$$
\begin{aligned}
H(F(\bar{X})+h^\beta \bar{W}_T,1) &= H^1 + h^\beta(< DF(\bar{X}),DF(\bar{X}) > +h^{2\beta}T)^{-1}\bar{W}_T, \\
H^1 &= (< DF(\bar{X}),DF(\bar{X}) > +h^{2\beta}T)^{-2} \\
&\quad \Big\{(< DF(\bar{X}),DF(\bar{X}) > +h^{2\beta}T)\int_0^T D_sF(\bar{X})\mathrm{d}W_s \\
&\quad +2\int_0^T\int_0^T D_sF(\bar{X})D_uF(\bar{X})D_uD_sF(\bar{X})\mathrm{d}u\mathrm{d}s\Big\},
\end{aligned}
$$

see e.g. Nualart [8, Exercise 2.1.1]. Then due to the independence of $W$ and $\bar{W}$ we have

$$E(1(F(\bar{X}) \geq x)H(F(\bar{X}) + h^{\beta}\bar{W}_T, 1)) = E(1(F(\bar{X}) \geq x)H^1).$$

From here the proof of the Theorem follows. □

Note that in the particular case that Condition (H1) is satisfied the above formulas simplify greatly with

$$\begin{aligned} H^1 &= (< DF(\bar{X}), DF(\bar{X}) >)^{-2}\Big\{(< DF(\bar{X}), DF(\bar{X}) >) \int_0^T D_s F(\bar{X}) dW_s \\ &\quad +2 \int_0^T \int_0^T D_s F(\bar{X}) D_u F(\bar{X}) D_u D_s F(\bar{X}) duds\Big\}. \end{aligned}$$

One should note that one can repeat the above arguments in order to carry more integration by parts if necessary.

# 7 Example with $F(X) = X_1$

Here we consider the case of uniform ellipticity in one dimension. That is, we assume that $b(x) \geq c > 0$ for all $x \in \mathbb{R}$. We also assume that $t_i = \frac{i}{N}$ for $i = 0, ..., N$, $k = 1$, and $F(X) = X_1$ for simplicity. In such a case it is known that the hypothesis (H) is satisfied and the variable $Y$ is

$$\begin{aligned} Y &= 1(X_1 \geq x)(\sum_{i=0}^{N-1} D_{t_{i+1}} \bar{X}_1 N^{-1} + N^{-2\beta})^{-2} \\ &\quad \Big\{(\sum_{i=0}^{N-1} D_{t_{i+1}} \bar{X}_1 N^{-1} + N^{-2\beta})(\sum_{i=0}^{N-1} D_{t_{i+1}} \bar{X}_1 (W_{t_{i+1}} - W_{t_i}) \\ &\quad + 2 \sum_{i,j=0}^{N-1} D_{t_{i+1}} \bar{X}_1 D_{t_{j+1}} \bar{X}_1 D_{t_{j+1}} D_{t_{i+1}} \bar{X}_1 N^{-2}\Big\}. \end{aligned}$$

As we have remarked above there are better ways of performing the simulation according to each specific situation. In particular there is a localized integration by parts formula that could be of use in certain situations. This formula is written as follows:

$$E(f'(F)G) = E(f(F)H_{\text{loc}}(F, G)) \quad \text{for} \quad m \geq 1,$$

where $H_{\text{loc}}(F, G) = \delta(G(\Delta_F^{\text{loc}})^{-1} g)$ and $g : \Omega \times [0, 1] \to \mathbb{R}$ is a smooth stochastic process such that $(\Delta_F^{\text{loc}})^{-1} = (< DF, g >)^{-1} \in \cap_{p>1} L_p(\Omega)$.

For example in the uniformly elliptic case it is known that $D_u X_1 > 0$ a.e. and that for a set of probability close to 1, $D_u \bar{X}_1 > 0$. This set is described as follows. Let $C := \|a'\|_\infty + \|b'\|_\infty$. Choose $N$, $M > 0$ satisfying $N \wedge M > 4C$. Then, on the set $L_M = \{\sup_{0 \leq k \leq m} |W_{t_{k+1}} - W_{t_k}| \leq 1/M\}$, $D_u \bar{X}_1 > 0$ for all $u \in [0, 1]$. Furthermore, $P(L_M^c) \leq C_q M^{-q}$ for any $q \geq 2$.

Using this fact one can simplify the above random variable $Y$ into

$$\begin{aligned} Y' &= 1(X_1 \geq x)(\sum_{i=0}^{N-1} D_{t_{i+1}} \bar{X}_1 N^{-1})^{-2} \\ & \left\{ W_1 \sum_{i=0}^{N-1} D_{t_{i+1}} \bar{X}_1 N^{-1} + \sum_{i,j=0}^{N-1} D_{t_{j+1}} D_{t_{i+1}} \bar{X}_1 N^{-2} \right\}. \end{aligned}$$

Here we have used as a localizing process $g = 1$ and the above definition is on the set $L_M$; otherwise $Y'$ is defined as 0.

# 8 Example with $F(X) = \int_0^T g(X_s)\mathrm{d}s$

In this section we consider as an example $F(X) = \int_0^T g(X_s)\mathrm{d}s$ for $d = 1$ and we assume that the coefficients $a$, $b$, $f$ and $g$ are smooth with bounded derivatives. This example is of importance in Finance in what is known as Asian options for which one takes $g(x) = x$. In this case one obtains that $\nabla_s F(X) = g'(X_s)$.

We check that the hypothesis (H) is satisfied in the following Lemma.

**Lemma 8.1** *Assume that $g'(X_u) \geq 0$ for almost all $u \in [0,T]$ and $g'(X_0) > 0$. Also assume that either $|b(x)| \geq c > 0$ for all $x \in \mathbb{R}$ or $a$ and $b$ are linear functions. Under the above conditions and definitions, $\Delta_{F(X)}^{-1} \in \cap_{p>1} L_p(\Omega)$.*

*Proof:* We will assume that $|b(x)| \geq c > 0$ for all $x \in \mathbb{R}$. The other case is done analogously. First one computes the Malliavin covariance matrix associated with $F(X)$. This calculation gives that

$$\Delta_{F(X)} = \int_0^T \int_s^T \int_s^T \nabla F_{u_1}(X) \mathcal{H}_{u_1} \mathcal{H}_s^{-1} b^2(X_s) \mathcal{H}_s^{-1} \mathcal{H}_{u_2} \nabla F_{u_2}(X) du_1 du_2 ds,$$

where $\mathcal{H}$ stands for the exponential associated with the derivative of the flow defined by $X$. That is, $\mathcal{H}$ is the unique $L^2$-adapted solution of

$$\mathcal{H}_t = 1 + \int_0^t a'(X_s)\mathcal{H}_s ds + \int_0^t b'(X_s)\mathcal{H}_s dW_s.$$

Using the hypothesis we have

$$\Delta_{F(X)} \geq C \inf_{0 \leq s \leq T} |\mathcal{H}_s^{-1}|^2 \int_0^T (\int_s^T g'(X_u)\mathcal{H}_u du)^2 ds$$

To the above one applies integration by parts twice in $s$ and Fubini's theorem to obtain

$$\Delta_{F(X)} \geq C \inf_{0 \leq s \leq T} |\mathcal{H}_s^{-1}|^2 \int_0^T 2g'(X_s)\mathcal{H}_s \int_0^s u g'(X_u)\mathcal{H}_u du ds. \tag{8.1}$$

Let $\delta > 0$ be such that $g'(x) > 0$ for $|x - X_0| < \delta$ Now define the following random time for fixed $\delta > 0$:

$$\tau = \inf\{s \geq 0; |X_u - X_0| \geq \frac{\delta}{2} \text{ or } |\mathcal{H}_s - 1| \geq 1/2$$
$$\text{or } |\mathcal{H}_s^{-1} - 1| \geq 1/2 \text{ or } |g'(X_u) - g'(X_0)| \geq g'(X_0)/2\}.$$

Recall that in order to prove that $\det(\Delta_{F(X)})^{-1} \in \cap_{p>1} L^p(\Omega)$ it is enough to prove that for all $p > 1$ there exists $\epsilon_0$ such that for all $\epsilon \leq \epsilon_0$ one has that

$$P^* := P(\Delta_{F(X)} \leq \epsilon) \leq \epsilon^p.$$

Approximating $\mathcal{H}$, $\mathcal{H}^{-1}$, $\nabla F(X)$ and $X$ by its values at 0 one has for $l \in (0, 1/3)$, by (8.1),

$$P^* \leq P(\tau \leq \epsilon^l) + P(C \inf_{0 \leq s \leq T} |\mathcal{H}_s^{-1}|^2 \int_0^\tau \int_0^s \frac{1}{8} g'(X_0)^2 u du ds < \epsilon, \ \tau > \epsilon^l).$$

Using the definition of $\tau$ and Chebyshev's inequality one obtains that $P(\tau \leq \epsilon^l) \leq C\epsilon^r$ for any $r > 0$. Similarly, one has

$$\begin{aligned} P(C \inf_{0 \leq s \leq T} |\mathcal{H}_s^{-1}|^2 g'(X_0)^2 \tau^3 < \epsilon, \ \tau > \epsilon^l) &\leq P(C g'(X_0)^2 \inf_{0 \leq s \leq T} |\mathcal{H}_s^{-1}|^2 < \epsilon^{1-3l}) \\ &\leq C\epsilon^{r(1-3l)} g'(X_0)^{-r} E(\sup_{0 \leq s \leq T} |\mathcal{H}_s|^{2r}). \end{aligned}$$

From here the result follows. □

A similar result can be achieved with the condition $g'(X_u) \leq 0$ for almost all $u \in [0, T]$ and $g'(X_0) < 0$. This example includes the cases of $g(x) = x^p$ for $p$ odd or any $p$ if $X$ is a positive stochastic process as in the case of linear stochastic differential equations that are of interest in finance.

Therefore all the previous results apply. In particular, one can also apply the principles used for the estimation of the density function to the estimation of the greeks associated to the corresponding European type options. The results are analogous to the ones exposed in section 6.

The following are some simulations of the densities obtained using the classical method and using the integration by parts formula.

# 9 Conclusion

The kind of result exposed here should allow efficient simulation of various different quantities that are expectations of nonsmooth functions in different situations. Although the theorems and examples have not been developed in full generality the emphasis is in methodology.

For example, one can develop the same theory for a general distribution function $T$ which will be approximated by the smooth function $T * \phi_h$ and similar results are obtained.

At the same time, one should also be cautious as to what extent these results hold. In general one requires smooth properties of the random variables involved in order to apply integration by parts of Malliavin Calculus. For example, the classical kernel density methods are usually applied when very little information is available about the random variables under consideration.

Asymptotically our results can be compared as follows in the case of approximations of density functions:

| Result type | Error size= $\alpha$ | Optimal step size or window size | Amount of calculations |
|---|---|---|---|
| Histograms for iid rv's | $n^{-1/3} = \alpha$ | $h = n^{-2/3} = \alpha^2$ | $\alpha^{-3}\alpha^{-2} = \alpha^{-5}$ |
| Corollary 5.4 | $n^{-1/3} = \alpha$ | $h = n^{-1/3} = \alpha$ | $\alpha^{-3}\alpha^{-1} = \alpha^{-4}$ |
| Theorem 6.2 | $n^{-1/2} = \alpha$ | $h = n^{-1/2} = \alpha$ | $\alpha^{-2}\alpha^{-1} = \alpha^{-3}$ |

The rate of convergence for histograms of iid samples is classical and can be found in e.g. Silverman [9]. According to Corollary 5.4 (for $\gamma = 1/2$) the window size is of the order $\sqrt{h}$ therefore giving a MISE (mean integrated squared error) of the order $(n\sqrt{h})^{-1} + O(h) + O(n^{-1})$ where $n$ is the sample size. The optimal choices for $n$ and $h$ for histograms of iid samples when the order of the error is specified is given in the first line of the above table.

The second line and third line are obtained from Corollary 5.4 and Theorem 6.2, respectively. It is clear that these improvements in rates are obtained due to the particular structure of stochastic differential equations. In general there are various counterexamples that show that the MISE rates are optimal.

Here the study is only asymptotical. The actual error will vary according to the situation and this requires a rather exact estimation of the constants that appear in the results presented here. Pilot studies show that in many cases the methods introduced improve on classical techniques.

But it is also clear from these studies that the constants involved in all the expansions in $h$ have a tendency to become bigger as the integration by parts formula is used. In further publications we will address other variance reduction techniques to deal with this problem.

# References


[1] V. Bally and D. Talay, The law of the Euler scheme for stochastic differential equations I: convergence rate of the distribution function, *Probab. Theory Rel. Fields* **104** (1996) 43-60.

[2] V. Bally and D. Talay, The law of the Euler scheme for stochastic differential equations: II. Convergence rate of the density, *Monte Carlo Methods Appl.* **2** (1996) 93-128.

[3] Y. Hu and S. Watanabe, Donsker delta functions and approximations of heat kernels by the time discretization method, *J. Math. Kyoto University* **36** (1996) 499-518.

[4] A. Kohatsu-Higa, High Itô-Taylor approximations to heat kernels, *J. Math. Kyoto University* **37** (1997) 129-150.

[5] P. E. Kloeden and E. Platen, *Numerical Solution of Stochastic Differential Equations*, Applications of Mathematics **23** (Springer, New York, 1992).

[6] S. Kusuoka and D. Stroock, Applications of the Malliavin Calculus II, *J. Fac. Sci. Univ. Tokyo Sec IA Math.* **32** (1985) 1-76.

[7] D. Nualart, *The Malliavin Calculus and Related Topics* (Springer Verlag, 1995).

[8] D. Nualart, Analysis on Wiener space and anticipating stochastic calculus, in: Ecole d'été de Probabilités de Saint-Flour XXV 1995, *Lecture Notes in Mathematics* **1690** (1998) 123–227.

[9] W. Silverman, *Density estimation* (Chapman Hall, London, 1986).

[10] D. Talay and L. Tubaro, Expansion of the global error for numerical schemes solving stochastic differential equations, *Stoch. Anal. Appl.* **8** (1990) 483-510 .